#### Additional Labelling Guideline

To label:

* train\_0200\_0250.tsv to train\_0350\_0400.tsv
* train\_1000\_1050.tsv to train\_2000\_2050.tsv

Notation:

1. *Examples are in italics*
2. Objects are marked with grey,
3. aspects are marked with red,
4. predicates with green.

**PRESENCE OF COMPARISON**

1. Pay attention to whether a **comparative adjective** compares the two objects or the object with the other version of itself or **is a part of an idiom**. In the latter two cases it shouldn’t be labelled:
   1. *cnn should know better to even bring that shit up they are no better than fox news .* - “should know better” is an idiom meaning that someone is doing something not sensible or smart. It’s not a comparison.
2. **Don’t label a comparative structure if the objects are not compared**, e.g. when the whole structure is questioned:
   1. *people seem to focus on the issue of bias , and this argument usually quickly devolves into a a stalemate between ideological camps , with people arguing about , for example , whether fox news is worse than cnn* - this is not a comparison but an indirect question, it gives no information on whether fox is really worse than cnn or not.
3. A sentence meaning “**Obj1 is the same as Obj2**”, it’s comparative. Possible comparative predicates for these cases are: **same**, **on par with**, **even, similar**.
   1. *postgresql 8 . 2 performance is at the same level , or better than oracle*
   2. *use postgresql as open source alternative , which has a feature - set similar to oracle database*
4. If **one of compared objects is not named explicitly**, don’t label such sentence:
   1. *the documentation sucks compared to either mysql or postgresql , but it ' s a million times better to work with than the oracle crap* - there’s some unknown object compared with mysql, postgresql, and oracle. We don’t know what object that is, so the sentence should be left unlabelled.

**ASPECT**

1. **Label aspects without prepositions**:
   1. *fox channels absolutely suck in quality compared to espn* - “in” (from “suck in quality”) shouldn’t be labelled.
   2. *fox news is easier to watch than cnn and that is saying something*
2. **Don’t label models of cars/phones and parts of objects as aspects**:
   1. *in fact , fox news unbalanced bias favoring romney makes them no better than cnn , msnbc etc* - “news” is a part of the channel name omitted here in order to get more examples of an object. It’s not the news which are compared, but channels in general, so “news” is not an aspect.
3. **Aspects are properties which are being compared**. They are usually (but not exclusively) expressed with a noun or a noun phrase:
   1. *hd picture on fox looks great tonight , about 4 million times better than the espn game that they tried to pass off as hd -* “looks” is not an aspect. What is compared is the *hd picture*.
   2. *fox regularly has a combined nightly audience , 5pm - midnite , greater than that of cnbc , msnbc and cnn , combined -* the comparison is “fox has greater audience than cnbc, msnbc, cnn”, not just “fox is greater than cnbc, msnbc, cnn”
4. In structures such as “**better at X**”, “easier to X”, “faster for X” and in general “JJS PREP NN”, **X should be an aspect**:
   1. *and , liberal bias or not , cnn is still the best for this type of coverage - better than the broadcast networks , much less fox* - “coverage” should be an aspect, the comparative phrase sounds as “better for <this type of> coverage”
5. If there is a structure “**easier to X and Y**”, both X and Y should be aspects:
   1. *postgresql can be easier to install and maintain than oracle*
6. **Label the minimal phrase as an aspect**. Don’t include adjectives or other non-essential details:
   1. *fox has by far the largest audience in the evenings -- larger than msnbc and cnn combined* - the aspect is “audience”, not “largest audience”, not “audience in the evenings”
7. **Aspect should be meaningful**. Ideally, it should be something which makes sense as a property of two objects even if we don’t see the sentence. If it doesn’t make sense on its own, it’s probably a bad aspect:
   1. *fox employs many liberals , whose numbers probably are greater than the number of conservatives employed by cnn and the other networks .* - Here the “number” could grammatically be an aspect, but it doesn’t make sense, we are comparing not just “numbers” for fox and cnn, but the number of liberals and conservatives working there. So “number” is not an aspect
8. **Don’t include aspect to the predicate**:
   1. *anybody wonder why fox news has much better ratings than cnn* - here the “ratings” is a feature being compared (an aspect), “better” is the realisation of comparison. “better ratings” shouldn’t be labelled as a multi-word predicate

**OBJECT**

1. If there are some **additional objects** (not labelled initially), label them as well:
   1. *the topic of fox news ... they aren ' t any worse than abc / nbc / cbs / msnbc / cnn / etc* - the initial objects are only “fox” and “cnn”, but other objects also are compared, so they should be labelled.
2. If the comparison is easily identifiable, but the **objects being compared are different from the ones initially labelled**, label the valid object and remove the labelling from the initial ones:
   1. Initial: *wow ... a cnn poll says dems are better off than the gop ... the same survey on fox said the opposite ( with a much larger margin*
   2. Should be: *wow ... a cnn poll says dems are better off than the gop ... the same survey on fox said the opposite ( with a much larger margin*

**PREDICATE**

1. If there is a **negation that changes the sense of a predicate**, it should be included into the predicate:
   1. *fox , or , as i call it , fake news is no better than the networks , cnn , etc* - here if we don’t include “no”, it will turn out that fox is better than cnn, whereas the sentence says the opposite.
   2. *oracle on a single system doesn ' t scale much better ( if even ) than postgresql on a single system* - “oracle is better than postgresql” opposes the sense of the sentence, so we should label “doesn’t” as a part of a predicate, because it’s essential for reconstructing the sense
2. A predicate can be expressed with a verb (if the verb has a meaning “is same/better/worse than”):
   1. *postgresql already beats oracle on cost*